64-13

MEMORANDUM
RM-3661-PR
APRIL 1964

SOME DIRECTIONS OF RESEARCH IN DYNAMIC PROGRAMMING

Richard Bellman

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM
RM-3661-PR
APRIL 1964

SOME DIRECTIONS OF RESEARCH IN DYNAMIC PROGRAMMING

Richard Bellman

This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA • 90406

PREFACE

In this Memorandum the author discusses some aspects and directions of research in the theory of dynamic programming, which is an important tool in the study of multistage decision processes.

SUMMARY

In this paper, the author indicates areas of research and some interesting and significant problems which arise in the attempt to use a certain functional equation as an effective algorithm in the study of multistage decision processes.

CONTENTS

SOME DIRECTIONS OF RESEARCH IN DYNAMIC PROGRAMMING

1. INTRODUCTION

Dynamic programming is a mathematical theory designed to implement (1) the study of multistage decision processes, and (2) the solution of problems which can be interpreted as arising from such processes. In a number of articles, and particularly in four books [1-4], we have explored some of the conceptual, analytic, and computational aspects of this new theory, and presented a variety of applications in engineering, economics, operations research, mathematical physics, and even in mathematics itself.

In the pages that follow, we wish to indicate some of the interesting, formidable, and significant problems that arise as soon as we attempt to use a characteristic functional equation such as

$$f_{n+1}(p) = \max_q [g(p,q) + f_n(T(p,q))] \tag{1.1}$$

as an effective algorithm for the numerical solution of questions in the areas named above.

As the reader will see, very little has been done so far, and there is unlimited opportunity for research in these new fields.

2. DIMENSIONALITY DIFFICULTIES

If p is a point in an n-dimensional space, $p = (p_1, p_2, \ldots, p_N)$, we face the problem of storing three functions of N variables, $f_n(p)$, $f_{n+1}(p)$, and $q_n(p)$, the maximizing value of q in (1.1), when we turn to this formula as a computational algorithm. Proceeding in a direct fashion, which is to say storing a function as the set of values it assumes, we see that if each component p_i is allowed k different values, then a total of $3 \times k^N$ values must be stored to determine the functions $f_n(p)$ in sequence, starting with

$$(2.1) \qquad f_1(p) = \max_q g(p,q).$$

It follows that if computing time is a factor, we cannot allow large values of k combined with values of $N \geq 3$. Rapid access storage at the moment is bounded by 32×10^3 words; with various simple devices we can push this to about 64×10^3 or at most 10^5.

Within the near future, we can contemplate rapid-access storage of 10^6, and probably within 25 years, following the usual curves of technological progress, we will have an upper limit of 10^9, with speeds about 10^3 or 10^6 faster than those current because of the uses of solid-state devices, miniaturization, and parallel circuitry.

Although these capabilities will trivialize many current problems, even these fabulous figures will not permit a routine approach to many other outstanding problems. Consider a situation where each component p_i is allowed to assume 100 different values. If $N = 4$, this leads to a total of 3×10^8 values; if $N = 6$, a total of 3×10^{12}.

The problem of storage of state variables is seen in even starker form, if we think in terms of processes with distributed parameters where the state variables are functions, or in terms of adaptive processes where the state variables are probability distributions.

It follows that it is essential, as in so many areas of classical analysis and mathematical physics, to think in terms of approximations. In the sections that follow, we shall discuss various types of approximate techniques and indicate the many new mathematical problems that arise in this fashion.

3. APPROXIMATION IN FUNCTION SPACE

The standard initial approach is to isolate those processes which possess simple analytic expressions for their optimal policies and return functions. An important class of processes of this nature are those where $g(p,q)$ is quadratic in its arguments and $T(p,q)$ is linear in p and q; see the work in [1-4] and [5-9]. Other important classes of processes

exist, however; for example, see [10], and the work on "optimal inventory" equations presented in [1] and [3], and the many results obtained in [11].

A fruitful approach would seem to be the study of the "inverse problem," i.e., given the optimal policy, determine all admissible return functions g(p,q) and transformations T(p,q), which lead to this policy. Some preliminary work has been done in [12] and [13].

Another classical direction is that of power series expansions in terms of state variables, time, or parameters appearing in the equation. Some preliminary results in connection with perturbation theory are given in [14].

A most important new direction, based upon the explicit solutions mentioned above, is that of quasi-linearization [15,16.17]. This theory is strongly connected with the concept of approximation in policy space, which we shall discuss below, but has its roots in classical analysis, particularly in functional inequalities (see [18], where reference is given to early work of Caplygin).

Finally, let us mention that the funtional-equation technique of dynamic programming applied to physical processes has produced the theory of invariant imbedding [19,20,21].

4. APPROXIMATION IN POLICY SPACE

In classical analysis, there exists only the technique of approximation in function space. Given an equation for an unknown function such as

$$f = T(f), \tag{4.1}$$

we generate a sequence $\{f_n\}$ by means of the relation

$$f_{n+1} = T(f_n). \tag{4.2}$$

Under suitable conditions, this sequence converges to a solution of (4.1).

Presented with the functional equation

$$f = \max_q T(f,q), \tag{4.3}$$

we can proceed as above. However, dynamic programming offers a new mode of approximation: approximation in policy space. In place of concentrating directly upon the return function f, we can focus upon the policy function q.

An advantage, of both analytic and computational significance, is that approximation in policy space automatically yields monotone convergence.

In this area, digital computers and mathematical experimentation can be extremely valuable. As far as the operational solution of control processes in the

engineering and industrial worlds is concerned, we know that there is a great premium on policies of simple type. The computational testing of large classes of simple policies in various types of multistage decision processes would yield valuable information concerning the need for more exact solutions and would furnish useful clues for further research.

Many years of application of Rayleigh-Ritz-Galerkin-type approximation procedures teach us that most functionals are fairly "flat." It would be important to make this more precise as far as optimization processes are concerned.

A new type of approximation process called "stochastic approximation" [22], [23] will play an important role in multistage decision processes. In any case, it is clear that there are many new ideas to be developed in these areas, ideas quite different from those of classical analysis.

5. APPROXIMATION IN INFORMATION PATTERN

Quite closely connected with the concept of approximation in policy space is the problem of the value of information concerning the state of the system. Suppose that we are given the values of only k of the components of p, or, more generally, k functions of the N components of p, how well can we make decisions? As $k \to N$, how rapidly does the optimal

return under partial information approach the optimal return from full information?

Put another way, how much does lack of information cost?

It is clear that there are many interesting classes of problems contained in this general format. What is currently called "Information Theory" is, as is shown in [24], only a very specialized version of a particular problem of this general type; see also [25].

Once again, mathematical experimentation with digital computers will be extremely helpful in guiding further research.

6. APPROXIMATION IN STRUCTURE SPACE

The fundamental objective of mathematical physics is the accurate approximation of the complex structures of reality by mathematical structures of simpler nature which are amenable to the analytic and computational techniques we currently possess. Dynamic programming and invariant imbedding represent new mathematical structures, differing in many ways from the classical approaches, designed to take advantage of the properties of the digital computer. There must be many such new approaches waiting for the discerning eye.

At the present time, there exists little work devoted to the structure of mathematical processes, and, in particular, to the concept of the degree of

approximation of one process by another. Considering how fruitful the problem of the approximation of a given function by a polynomial has been, it is clear that many significant problems exist in this area.

7. DIFFERENTIAL APPROXIMATION

As a step in the preceding direction, let us consider the technique of differential approximation [26]. The approximation of a function $f(t)$ on the interval $-\infty < a \leq t < b < \infty$ by a polynomial

$$(7.1) \qquad p_n(t) = \sum_{k=0}^{n} a_k t^k$$

is equivalent to the approximation of $f(t)$ by means of a solution of the linear differential equation

$$(7.2) \qquad \frac{d^{n+1}u}{dt^{n+1}} = 0.$$

Similarly, to approximate to $f(t)$ by means of an exponential polynomial of the form

$$(7.3) \qquad p_n(t) = \sum_{k=0}^{n} a_k e^{\lambda_k t}$$

is to approximate to $f(t)$ by means of a solution of the general linear differential equation with constant coefficients :

$$(7.4)\qquad \frac{d^{n+1}u}{dt^{n+1}} + b_1 \frac{d^n u}{dt^n} + \cdots + b_n u = 0.$$

In place of the usual approach to approximation problems of this nature, let us proceed in the following way. Determine the coefficients b_i in (7.4) by the condition that the quadratic form

$$(7.5)\qquad \int_0^T \left(\frac{d^{n+1}u}{dt^{n+1}} + b_1 \frac{d^n u}{dt^n} + \cdots + b_n u \right)^2 dt$$

is minimized. If u is determined as the solution of a nonlinear differential equation

$$(7.6)\qquad \frac{d^k u}{dt^k} = g\left(u, \frac{du}{dt}, \ldots, \frac{d^{k-1}u}{dt^{k-1}}\right),$$

the numerical work is easily carried out; see [26].

Problems that arise are those of degree of approximation and type of approximation to use. In the approach sketched above, $2n + 2$ parameters are required to determine $u(t)$, the $n + 1$ coefficients b_i and $n + 1$ initial values. Can we obtain superior approximation by a different allocation of parameters? For example, might it be better to use a nonlinear differential of lower order, or variable coefficients? Once again, some mathematical experimentation will yield suggestions concerning further research.

8. SEARCH PROCESSES

So far, we have concentrated upon the approximation to the return function or the policy function as a fundamental problem to be overcome in the use of (1.1). An equally important problem as far as cutting down on computing time is concerned is that of obtaining the maximum over q. If $\underset{\sim}{q}$ is multidimensional, straight-forward enumerative search will consume far too much time.

The basic problem is that of utilizing the structural features of the process to accelerate the search process. As such, we see that we are verging upon pattern recognition processes, and, indeed, the two are closely related.

A brief discussion of some recent work is contained in [3]. In general, however, little has been done in this new field, and there is ample room for extensive research.

9. SCHEDULING PROBLEMS AND COMPUTERS

Ideally, what is desired is a self-organizing computer which will rearrange its components and scheduling so as to solve a particular problem in a most efficient fashion. Some work has been done in this connection (see [27]).

Leaving aside an optimal reorganization, even a parallel-operation computer would be an enormous advance.

In any case, vast opportunities exist in these directions to bring about orders-of-magnitude reduction in computing time. An advance of this type would be equivalent to a mathematical advance of an order of magnitude.

10. CONCLUSION

We have attempted to discuss areas of research, rather than particular problems. The reader interested in specific questions may refer to the books [1-4].

REFERENCES

1. Bellman, R., Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

2. ———, Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, New Jersey, 1961.

3. Bellman, R., and S. Dreyfus, Applied Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1962.

4. Bellman, R., and R. Kalaba, Dynamic Programming and Modern Control Theory, McGraw-Hill Book Company, Inc., New York, to appear.

5. Beckwith, R., Analytic and Computational Aspects of Dynamic Programming Processes of High Dimension, Ph.D. Thesis, Purdue University, 1959.

6. Freimer, M., A Dynamic Programming Approach to Adaptive Control Processes, Lincoln Laboratory, No. 54-2, 1959.

7. Adorno, D., The Asymptotic Theory of Control Systems—I: Stochastic and Deterministic Processes, Jet Propulsion Lab., Technical Release 34-73, June 30, 1960.

8. Kalman, R., New Methods and Results in Linear Prediction And Filtering Theory, RIAS, 1962.

9. Bellman, R., Introduction to Matrix Analysis, McGraw-Hill Book Company, Inc., New York, 1960.

10. Windeknecht, T. G., "Optimal Stabilization of Rigid Body Attitude," Journal of Mathematical Analysis and Applications, Vol. 6, No. 2, 1963, pp. 325-335.

11. Arrow, K. J., S. Karlin, and H. Scarf, Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, California, 1958.

12. Bellman, R., and R. Kalaba, An Inverse Problem in Dynamic Programming and Automatic Control, The RAND Corporation, RM-3592-PR, April 1963.

13. Kalman, R., On the Inverse Problem, RIAS, 1963.

14. Bellman, R., A Note on Dynamic Programming and Perturbation Theory, The RAND Corporation, RM-3169-PR, June 1962.

15. Aoki, M., "On a Successive Approximation Technique in Solving Some Control System Optimization Problems," Journal of Mathematical Analysis and Applications, Vol. 5, No. 3, 1962, pp. 418-434.

16. Bellman, R., "Functional Equations in the Theory of Dynamic Programming—V: Positivity and Quasi-linearity," Proceedings of the National Academy of Sciences, USA, Vol. 41, 1955, pp. 743–746.

17. Kalaba, R., "On Nonlinear Differential Equations, the Maximum Operation and Monotone Convergence," Journal of Mathematics and Mechanics, Vol. 8, 1959, pp. 519–573.

18. Beckenbach, E. F., and R. Bellman, Inequalities, Ergebnisse der Math., Springer, Berlin, 1961.

19. Bellman, R., R. Kalaba, and G. M. Wing, "Invariant Imbedding and Mathematical Physics—I: Particle Processes," Journal of Mathematical Physics, Vol. 1, 1960, pp. 280–308.

20. Bellman, R., and R. Kalaba, "Invariant Imbedding, Wave Propagation, and the WKB Approximation," Proceedings of the National Academy of Sciences, USA, Vol. 44, 1958, pp. 317–319.

21. Bellman, R., R. Kalaba, and M. Prestrud, Invariant Imbedding and Radiative Transfer in Slabs of Finite Thickness, The RAND Corporation, R-388-ARPA, 1962.

22. Robbins, H., and S. Monro, "A Stochastic Approximation Model," Annals of Mathematical Statistics, Vol. 22, 1951, pp. 400–407.

23. Dvoretzky, A., "On Stochastic Approximation," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley, California, 1956, pp. 39–56.

24. Bellman, R., and R. Kalaba, "On the Role of Dynamic Programming in Statistical Communication Theory," IRE Transactions of Professional Group on Information Theory, Vol. IT-3, 1957, pp. 197–203.

25. Marshak, J., Remarks on the Economics of Information, Cowles Foundation Discussion Paper No. 70, April 1959.

26. Bellman, R., R. Kalaba, and B. Kotkin, Differential Approximation Applied to the Solution of Convolution Equations, The RAND Corporation, RM-3601-NIH, May 1963.

27. Aoki, M., and G. Estrin, The Fixed-Plus-Variable Computer System in Dynamic Programming Formulation of Control System Optimization Problems. Part I, Report No. 60-66, Engineering Department, University of California, Los Angeles.